REMARKS

The Examiner has rejected claims 1-31. Claims 1-31 are pending for examination with claims 1,8,18, and 22 being independent claims.

The Examiner has rejected Claim 1 and 18 under 35 U.S.C. §102(b) as being anticipated by U.S. Patent No. WO 99/62007 to Fayyad et al. ("Fayyad").

Applicants have amended Claim 1 to call for:
"c) performing <u>clustering of data records in two phases</u> including a first phase and a second phase, <u>the first phase clustering the data records over a discrete attribute space,</u> and the <u>second phase clustering continuous attributes</u> the data to produce an intermediate set of data clusters" (underlining added for emphasis)

Applicants have amended Claim 18 to call for:
"d) said computer including a stored program for
i) grouping together data records from the database which have specified discrete attribute configurations;
ii) <u>a first clustering of data records having the same or a similar specified discrete attribute configuration</u>
iii) <u>a second clustering data records based on the continuous attributes</u>; and
iv) merging together the first clustering and the second clustering to produce a clustering model." (underlining added for emphasis)

Applicants have amended Claim 22 to call for:
"b) <u>performing a first clustering of</u> data records from the database which have specified <u>discrete attribute</u> configurations;
c) <u>performing a second clustering of the</u> data records having the same or similar specified discrete attribute configuration based on the <u>continuous attributes</u> to produce an intermediate set of data clusters"
(underlining added for emphasis)

As such, Applicants submit that Claim 1, 18 and 22 are not anticipated by Fayyad under 35 U.S.C. §102(b).

The present invention provides for:

> "In accordance with the present invention, the clustering model 15 is arrived at in two phases. A cluster structure over the discrete attribute space is first performed using methods similar to methods for identifying frequent itemsets in data. Known frequent itemset identification algorithms are efficient in dealing with 1000 - 100,000s of attributes. The present invention uses similar methods to locate discrete attribute cluster structure. Once this cluster structure is determined, structure over the continuous attributes of data is identified using one of the many methods currently available for clustering continuous attribute data. " (8:1-8) (underlining added for emphasis).

Fayyad, on the other hand provides for

> "Mixed Data Clustering Model Now assume that instead of including only data records with discrete non- ordered data, the records read from the database 12 have discrete data like color and ordered (continuous) attributes such as a salary field and an age field. These additional fields are continuous and it makes sense to take the mean and covariance, etc. of the values for these additional fields. For each of the 3 clusters being modeled, one can assign a Gaussian (having a mean and covariance matrix) to the income and age attributes and calculate contributions to each cluster for each data record based upon its attribute values.

> Now again consider the records from Table 1. In addition to the previously discussed three attributes of'color','style'and'sex', each record has the additional attributes of'income'and'age'. These mixed attribute records are listed below in Table 3. Note, the female that purchased the blue sedan (RecordId #2) is now further classified with the information that she has an income of 46K and an age of 47 years.

Table 3 RecordID Color Style Sex Income Age 1 yellow sedan male 24K 32yrs 2 blue sedan female 46K 47 3 green sedan male 82K 66 4 white truck male 40K 30 5 yellow sport female 38K 39 <u>For each of the records of Table 3 the data mining engine 12 must compute the probability of membership of each data record in each of the three clusters</u>. Suppose, in the general case, the discrete attributes are labeled"DiscAtt#1", "DiscAtt#2",...,"DiscAtt#d"and let the remaining continuous attributes make up a numerical vector x. The notation for determining this probability is: Prob (record! cluster #) = p (DiscAtt#1 I cluster #) *p (DiscAtt#2 I cluster #) *... *p (DiscAtt#d cluster #) * p (x 1 4, F, of cluster #). Here p (DiscAttr#j | cluster #) is computed by looking up the stored probability of DiscAttr#j in the given cluster (i. e. reading the current probability from the attribute/value probability table associated with this cluster). p (x | >, E of cluster #) is calculated by computing the value of x under a normal distribution with mean µ and covariance matrix E:" (11:42 through 12:32) (underlining added for emphasis)

Accordingly, Applicants submit that Claim 1, 18 and 22 are not anticipated by Fayyad under 35 U.S.C. §102(b).

Claims 2-7 are dependent on Claim 1. As such, Claims 2-7 are believed allowable based upon Claim 1.

Claims 19-21 are dependent on Claim 18. As such, Claims 19-21 are believed allowable based upon Claim 18.

Claims 23-31 are dependent on Claim 22. As such, Claims 23-31 are believed allowable based upon Claim 22.

The Examiner has rejected Claim 8 under 35 U.S.C. §102(b) as being anticipated by U.S. Patent No. WO 99/62007 to Fayyad et al. ("Fayyad").

Applicants have amended Claim 8 to call for:
"b) <u>performing a first discrete cluster by counting data records from the database which have the same discrete attribute</u> configuration and identifying a first set of

configurations wherein the number of data records of each configuration of said first set of configurations exceeds a threshold number of data records;

...

d) <u>performing a second continuous clustering</u> of the subset of records contained within at least some of the first set of configurations <u>based on the continuous data attributes</u> of records contained within that first set of configurations to produce a clustering model."
(underlining added for emphasis)

As such, Applicants submit that Claim 8 is not anticipated by Fayyad under 35 U.S.C. §102(b).

The present invention provides for:
"In accordance with the present invention, <u>the clustering model 15 is arrived at in two phases</u>. A cluster structure <u>over the discrete attribute space</u> is first performed using methods similar to methods for identifying frequent itemsets in data. Known frequent itemset identification algorithms are efficient in dealing with 1000 – 100,000s of attributes. The present invention uses similar methods to locate discrete attribute cluster structure. <u>Once this cluster structure is determined, structure over the continuous attributes of data is identified</u> using one of the many methods currently available for clustering continuous attribute data. " (8:1–8) (underlining added for emphasis).

Fayyad, on the other hand provides for:
"A data structure for the results or output model of the analysis for the ordered attributes is depicted in Figure 8D. This model includes K data structures for each cluster. <u>Each cluster is defined by</u> 1) a vector'Sum'representing the sum of each of the database records for each of the ordered or <u>continuous attributes</u> or dimensions (n = number of continuous attributes), 2) a vector'Sumsq'representing the sum of the continuous attributes squared, 3) a floating point value'M'counting the number of data records contained in or belonging to the

corresponding cluster, and 4) <u>an attribute/value</u>
<u>probability table</u> such as the table depicted in Figure 9A,
<u>summarizing the discrete attributes</u> (d = number of
discrete attributes). " (15:19-27) (underlining added for
emphasis)

Accordingly, Applicants submit that Claim 8 is not anticipated by Fayyad under
35 U.S.C. §102(b).

Claims 9-17 are dependent on Claim 8. As such, Claims 9-17 are believed
allowable based upon Claim 8.

CONCLUSION

Accordingly, in view of the above amendment and remarks it is submitted that the claims are patentably distinct over the prior art and that all the rejections to the claims have been overcome. Reconsideration and reexamination of the above Application is requested. Based on the foregoing, Applicants respectfully requests that the pending claims be allowed, and that a timely Notice of Allowance be issued in this case. If the Examiner believes, after this amendment, that the application is not in condition for allowance, the Examiner is requested to call the Applicant's attorney at the telephone number listed below.

If this response is not considered timely filed and if a request for an extension of time is otherwise absent, Applicants hereby request any necessary extension of time. If there is a fee occasioned by this response, including an extension fee that is not covered by an enclosed check please charge any deficiency to Deposit Account No. 50-0463.

Respectfully submitted,

Microsoft Corporation

Date: <u>October 17, 2005</u>

By: _____

Microsoft Corporation
One Microsoft Way
Redmond, WA 98052-6399

Paul B. Heynssens, Reg. No.: 47,648
Attorney for Applicants
Direct telephone (425) 707-3913

<table>
<tr><td colspan="2" align="center">CERTIFICATE OF MAILING OR TRANSMISSION
UNDER 37 C.F.R. § 1.8(a)</td></tr>
<tr><td colspan="2">I hereby certify that this correspondence is being:

☒ deposited with the United States Postal Service on the date shown below with sufficient postage as first class mail in an envelope addressed to: Mail Stop AF, Commissioner for Patents, P. O. Box 1450, Alexandria, VA 22313-1450.</td></tr>
<tr><td><u>October 17, 2005</u>
Date</td><td>_____
Signature

<u>Noemi Tovar</u>
Type or Print Name</td></tr>
</table>

Microsoft Corporation
MS 163193.01